**Used Cars Dataset workshop** based on the topics we’ve covered

### **🔹1. Data Cleaning Challenges**

**Task:** The dataset contains missing values and inconsistencies. Clean it up!

* Find and list missing values for each column.
* Fill missing values logically (e.g., median price for missing prices).
* Remove duplicates (if any).
* Convert categorical features (like fuel type, transmission) into consistent formatting.
* Detect and remove outliers in car prices using **Z-score or IQR**.

### **🔹 2. Pandas Data Manipulation Challenges**

### **Task:** Extract insights using Pandas.

* Calculate the **average mileage** per car brand.
* Find how many cars have a price higher than **1.5 times the median price**.
* Extract **year of manufacture** from the dataset and group by decade.
* Create a **new feature**: price\_per\_mile = price / mileage.

### **🔹 3. Data Visualization Challenges**

**Task:** Use **Matplotlib/Seaborn** to visualize insights.

* Plot a **histogram** of car prices.
* Create a **scatter plot** of price vs. mileage (what’s the trend?).
* Show a **bar chart** of the top 10 most common car brands.
* Compare the **average price of petrol vs. diesel vs. electric cars** using a box plot.

### **🔹 4. Exploratory Data Analysis (EDA) Challenges**

**Task:** Identify interesting trends.

* What is the **relationship between price and age** of the car?
* Do electric cars have significantly different pricing than petrol cars?
* Which **brands hold their value the best** (smallest price drop over time)?
* Are there seasonal trends in **car pricing** (if dates are available)?

### **🔹 5. Machine Learning Challenges**

**Task:**

* Preprocess the data (handle missing values, categorical encoding).

lan for term 2

Preprocessing Review# Tom - week 2

Excel

Power BI week 4 (Techvision) max

R (Izzy)

Git / Github - week 3

R / RStudio  
Neural Networks (Keras) (Techvision) - Week 10

NoSQL (MongoDB)\*\*

Data Science in Finance\*\* (Collab) ( More SQL)